

order to compute the TED score. Inference timing results for all experiments were obtained from the same machine on a single core with AMD EPYC 7763 CPU @2.45 GHz.

### 5.1 Hyper Parameter Optimization

We have chosen the PubTabNet data set to perform HPO, since it includes a highly diverse set of tables. Also we report TED scores separately for simple and complex tables (tables with cell spans). Results are presented in Table. 1. It is evident that with OTSL, our model achieves the same TED score and slightly better mAP scores in comparison to HTML. However OTSL yields a *2x speed up* in the inference runtime over HTML.

**Table 1.** HPO performed in OTSL and HTML representation on the same transformer-based TableFormer [9] architecture, trained only on PubTabNet [22]. Effects of reducing the # of layers in encoder and decoder stages of the model show that smaller models trained on OTSL perform better, especially in recognizing complex table structures, and maintain a much higher mAP score than the HTML counterpart.

| #<br>enc-layers | #<br>dec-layers | Language | TEDs   |         |              | mAP<br>(0.75) | Inference<br>time (secs) |
|-----------------|-----------------|----------|--------|---------|--------------|---------------|--------------------------|
|                 |                 |          | simple | complex | all          |               |                          |
| 6               | 6               | OTSL     | 0.965  | 0.934   | 0.955        | <b>0.88</b>   | <b>2.73</b>              |
|                 |                 | HTML     | 0.969  | 0.927   | 0.955        | 0.857         | 5.39                     |
| 4               | 4               | OTSL     | 0.938  | 0.904   | 0.927        | <b>0.853</b>  | <b>1.97</b>              |
|                 |                 | HTML     | 0.952  | 0.909   | <b>0.938</b> | 0.843         | 3.77                     |
| 2               | 4               | OTSL     | 0.923  | 0.897   | 0.915        | <b>0.859</b>  | <b>1.91</b>              |
|                 |                 | HTML     | 0.945  | 0.901   | <b>0.931</b> | 0.834         | 3.81                     |
| 4               | 2               | OTSL     | 0.952  | 0.92    | <b>0.942</b> | <b>0.857</b>  | <b>1.22</b>              |
|                 |                 | HTML     | 0.944  | 0.903   | 0.931        | 0.824         | 2                        |

### 5.2 Quantitative Results

We picked the model parameter configuration that produced the best prediction quality (enc=6, dec=6, heads=8) with PubTabNet alone, then independently trained and evaluated it on three publicly available data sets: PubTabNet (395k samples), FinTabNet (113k samples) and PubTables-1M (about 1M samples). Performance results are presented in Table. 2. It is clearly evident that the model trained on OTSL outperforms HTML across the board, keeping high TEDs and mAP scores even on difficult financial tables (FinTabNet) that contain sparse and large tables.

Additionally, the results show that OTSL has an advantage over HTML when applied on a bigger data set like PubTables-1M and achieves significantly improved scores. Finally, OTSL achieves faster inference due to fewer decoding steps which is a result of the reduced sequence representation.